Causal Inference in Law: An Epidemiological Perspective

*Bob Siegerink, Wouter den Hollander, Maurice Zeegers and Rutger Middelburg**

I. Introduction

Causal inference lies at the heart of many legal questions. Yet in the context of complicated disease litigation, in particular, the causal inquiry is beset with difficulties due to gaps in scientific knowledge concerning the precise biological processes underlying such diseases. Civil courts across the globe, faced with increased litigation on such matters, struggle to adhere to their judicial fact-finding and decision-making role in the face of such scientific uncertainty. An important difficulty in drawing evidentially sound causal inferences is the binary format of the traditional legal test for factual causation, being the 'but for' test, which is based on the *condicio-sine-qua-non* principle.[1] To the question 'would the damage have occurred in the absence of the defendant's wrongful behaviour' the 'but for' test requires a simple yes or no answer. This is increasingly deemed unsatisfactory in cases in which, given the state of science, true causation cannot possibly be determined with certainty. Given the general rule that the burden of proof in principle lies with the claimant, the 'but for' test passes on the uncertainty to the claimant entirely. Such is not only felt to be at odds with fairness, but is also unsatisfactorily from an epidemiological perspective, given the binary format of the 'but for' test on the one hand and the fact that most diseases are multi-causal and cannot be ascribed to a single factor only on the other hand.

In this article, we will elaborate this epidemiological perspective and from that perspective discuss the problem of causal inference in law in general and scrutinize one new legal concept dealing with this problem in particular. This is the concept of the so-called proportional liability, as accepted by the Dutch Supreme Court in the Nefalit-case. The Supreme Court agreed with the lower courts, assuming liability of employer Nefalit, in proportion to the reasoned estimation of the chance that the lung cancer Karamus suffered from was caused by asbestos exposure during the work for his employer Nefalit (55%). We will argue that although such proportional liability adheres to the epidemiological concept of multi-causality, and in that respect, is not without merit, epidemiological measurements *on a population level* should not be taken to calculate the probability that the employers' wrongful conduct has actually caused the disease in an *individual*. We propose a different approach in two stages, making proportional liability more truly proportional to the defendant's relative contribution in the known causal mechanism underlying the damage in question and, by that, more fair for both parties, even though our approach is not flawless either.

We will set out some important concepts from the field of epidemiology with respect to causal inference first. A thorough understanding of these concepts will help to further strengthen and inform legal principles of causation. Epidemiology, where probabilistic concepts are applied to address causal questions in individuals, could in particular aid in the understanding of multi-causality and its possible links to proportional liability as a legal concept. Epidemiology studies the distribution and determinants of disease frequency in human populations. It contrasts with daily medical practice which focuses on individuals. We will elaborate the difference in concepts of causal inference between groups and individuals, with a link to the *condicio-sine-qua-non* principle and the concept of multi-causality. We will then discuss

* Bob Siegerink, Phd: Center for Stroke Research, Charité, Universitätsmedizin Berlin, Berlin, Germany and Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Nehterlands; P. Wouter Hollander, LLM MA: Institute for Private Law, Leiden University, Leiden, the Netherlands; Maurice P. Zeegers, PhD: Maastricht University, School of Nutrition, Toxicology and Metabolism & Maastricht Forensic Institute, Maastricht, the Netherlands; Rutger A. Middelburg, PhD: Center for Clinical Transfusion Research, Sanquin Research, Leiden, the Netherlands, and Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. The authors would like to acknowledge AG Castermans and Ruben de Graaff (both Institute for Private Law, Leiden University) for their useful comments.

1 See for instance Deakin, S., Johnston, A., Markesinis, B. *Markesinis and Deakin's Tort Law* (7th ed.), Oxford: Clarendon Press 2013, pp. 218-256.

how causation can be quantified in a single number and how these numbers compare to the legal concept of proportional liability as accepted in the Nefalit-case. Ultimately, we will try to reconcile the impossibility to know the exact causal mechanism of a disease in an individual to the *condicio-sine-qua-non* principle and the application of proportional liability to come to a fair reimbursement of damages in complex disease litigation.

II. Causal Inference in Medicine and Epidemiology

In clinical medicine, doctors are confronted with questions of causality on a daily basis. Will this medical treatment cause the cure of a patient? And will the benefits outweigh the side-effects caused by this treatment? For example, when a patient suffers from an ischaemic stroke caused by a blood clot in the brain that is preventing the flow of oxygenated blood, a decision can be made to start treatment targeted to resolve the blood clot and restore blood flow. This treatment is called thrombolysis and restores blood flow in 43% of treated cases.[2] However, thrombolysis also causes bleedings, which in itself can also be a cause of morbidity and mortality.[3] Thrombolysis can only be applied in the first 3-4,5 hours after the onset of symptoms, because only in this time period the benefits of treatment, which declines over time, outweigh the negative consequences of this treatment on a population level.

Treating a patient is not restricted to addressing the acute symptoms of a certain cause, but also includes the removal of possible causes to prevent a possible recurrence of the disease. For example, the physician confronted with a patient suffering from an ischaemic stroke will not only apply thrombolysis, but will also target the smoking habit and the increased cholesterol levels of that particular patient to prevent another case of ischaemic stroke in the long run. The decision to target these risk factors is based on studies that *on a population level* these factors area cause of the disease. Targeting these risk factors in an *individual* is therefore thought to lower the risk of recurrence.[4,5] But how can the physician, based on epidemiological studies, be certain that the smoking habit and high cholesterol levels were causal in the mechanism leading to the ischaemic stroke in *this* particular patient? The unsettling answer is that he is not certain, neither can he ever be.

III. The Counterfactual Ideal

Theoretically, we can only be certain on the causal nature of a risk factor if we observe the outcome when the patient is exposed to this risk factor and compare that to the situation when we go back in time, and see what happens if the patient is unexposed, but all other factors are kept constant.[6] Because this hypothetical situation is contrary to fact, this concept is sometimes referred to as the counterfactual or potential outcome model.[7,8] If we could go back in time, and manipulate only one certain factor we could determine in each individual patient whether an individual risk factor was indeed a cause of the observed disease.

This counterfactual model is comparable to the *condicio-sine-qua-non*-test in law. The risk that describes this relationship between exposure and disease for one individual is binary, being 1 (for the disease is caused by the exposure) or 0 (for the disease is not caused by the exposure). However, since the counterfactual outcome cannot be observed, we cannot determine the causal mechanism in an individual. The counterfactual ideal can be approached, though, in the comparison of different populations under certain conditions. For example, if two groups are similar except for the presence of the risk factor of interest, a difference in disease frequency can be ascribed to the sole difference between these groups,

2 Joung, H. Rha, & Saver, L.J., 'The Impact of Recanalization on Ischaemic Stroke Outcome: a Meta-Analysis', *Stroke* 38 (2007), pp. 967–73.

3 Lansberg, M.G. *et al.*, 'Antithrombotic and Thrombolytic Therapy for Ischaemic Stroke: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines', *Chest* 141 (2012), e601S–36S.

4 Goya Wannamethee, S. *et al.*, 'Smoking Cessation and the Risk of Stroke in Middle-Aged Men', *JAMA* 274 (1995), pp. 155–60.

5 Milionis, H.J. *et al.*, 'Statin Therapy after First Stroke Reduces 10-year Stroke Recurrence and Improves Survival', *Neurology* 72 (2009), pp. 1816-22.

6 For more background reading on the theory of causation, please refer to: Pearl, J., Causality: Models, Reasoning, and Inference, Cambridge University Press, 2000; 2nd edition, 2009.

7 Rothman, K.J. *et al.*, 'Causation and Causal Inference in Epidemiology', *Am. J. Public Health* 95 (2005) Suppl. 1, pp. S144-50.

8 Rothman, K.J., Greenland, S., Lash, T.L., *Modern Epidemiology* (third revised edition), Lippincott Williams & Wilkins 2008.

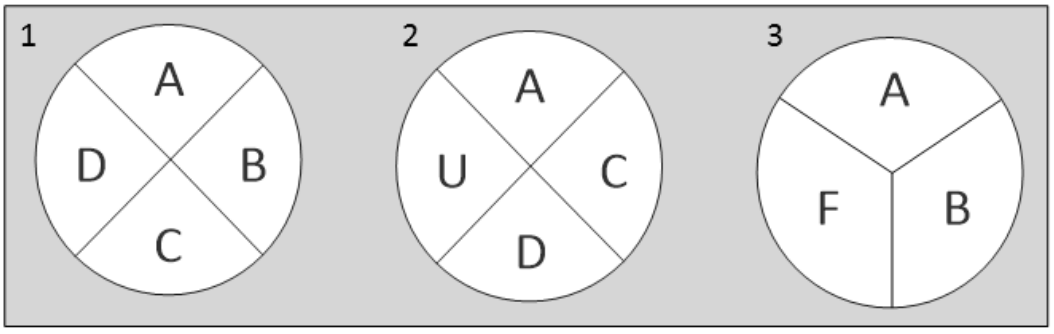

Figure 1 - Three sufficient causes

being the risk factor of interest. This comparison does not allow to establish the causal mechanism within an individual. However, these group comparisons do allow us to estimate the causal relationship between the exposure and the outcome to be quantified in terms of probability.

IV. From One Cause to the concept of Multi-causality

Before we describe how causal relationships can be quantified, we first have to focus on the definition of a causal mechanism. Often, the cause in a causal mechanism is thought to be a single factor in a cause-consequence sequence. However, a consequence can have multiple causes: several factors, as a combination, cause an effect. This concept is known as multi-causality and is important in epidemiological thinking on causality, for it provides a way to think about causal mechanisms instead of single cause-consequence sequences. To avoid confusion, multi-causality should be distinguished from the situation in which a single factor, such as smoking, can cause different diseases.

In epidemiological theory, the concept of multi-causality has gained ground since it was formalized by K.J. Rothman in 1976.[9] The concept distinguishes and describes the implications of *necessary*, *sufficient* and *component causes* and is further explained with the use of figure 1.

Let there be three possible causal mechanisms that lead to a certain disease. Figure 1 depicts these three causal mechanisms as three *sufficient causes*, all comprising multiple *component causes*. In this example we assume that these three sufficient mechanisms are the only three possible causal mechanisms that lead to an event, which can be a disease, injury or anything similar. The frequency of the disease normally is a direct function of the frequency of – a combination of – the different component causes. We also assume that all component causes are equally present in the population and that the presence of each component cause is independent from the others (i.e. no confounding causes, see below).

Important to note is that sometimes component causes are present in all sufficient causes, making them *necessary component causes* (A in our example). In theory, removal of a necessary component cause from the population will lead to complete eradication of the disease. It is not necessary for all sufficient causes to have an equal number of component causes, nor is it needed to name all component causes in detail. A component cause can even be unknown (often depicted as 'U', as is done in the middle sufficient cause in figure 1).

A component cause can be present long before the sufficient cause is completed. For example, a genetic variation in a certain gene is present from before birth, but other component causes are needed to complete a sufficient cause. The completion of a sufficient cause equals the biologic onset of the disease, which is not necessarily the time of diagnosis. These concepts are illustrated in an example where genetic variations are part of the causal mechanism leading to ischaemic stroke: genetic variations in the APOE gene are known to cause blood cholesterol levels to rise. These genetic variations are present from before birth, but this small increase in blood cholesterol alone is in itself insufficient and additional car-

9 Rothman, K.J., 'Causes', *Am. J. Epidemiol* 104 (1976), pp, 587-92.

diovascular risk factors are needed in order to cause an ischaemic stroke. Together with such factors (e.g. smoking), increased blood cholesterol might result in an atherosclerotic plaque. Sometimes, these plaques rupture and a thrombus is formed, which subsequently blocks the flow of blood to the brain. Also, the moment of diagnosis of the ischaemic stroke or even the first symptoms can be hours later than the actual blockage of the cerebral artery.

Since component causes accumulate over time, the incidence of many diseases rises sharply with age. The time between the first presence of a component cause and the completion of the sufficient cause is referred to as the induction time. In our example, the alphabetical order of the components refers to the order in which they occur. It is important to note that the length of the induction does not necessarily reduce the importance of a particular component cause. The component cause that completes the sufficient cause has an induction time of zero and is therefore easily identified as a cause. Component causes with little to no induction time are in layman's terms for that reason sometimes erroneously referred to as *the* cause of the disease.

Nonetheless, the order of component causes is of importance: a person who is only exposed to component causes A and B has no sufficient cause. Subsequent exposure to the two components causes C and D will complete a sufficient cause. When this person is not exposed to C or D, he will not develop the disease at that particular point in time. However, when this same person at a later moment is exposed to component cause F, a sufficient cause has formed and the person still develops the disease, albeit somewhat later in time.

The sufficient cause model adheres to the counterfactual ideal. When we consider the sufficient cause 1 depicted in figure 1, we can see that A, B, C and D are the component causes for this particular sufficient cause. If we think of the counterfactual situation that this particular individual was not exposed to component cause A and all other things equal, this disease would not have occurred. The same goes for the common causes B, C and D. We can even broaden our view and see what happens with the whole population: if necessary component cause A were to be eliminated from the population, 100% of all sufficient causes cannot be formed anymore and the disease would have been eradicated from the population.

We can also see that 2/3 of the sufficient causes comprise component cause D. Removing D from our population would however not necessarily reduce the number of diseased in our population by this same number. After all, persons with sufficient cause 1 are now only exposed to component cause A, B and C and therefore still at risk of developing the disease for example when exposed to component cause F later in time. If in the extreme case each individual exposed to component cause D is also, at some later time, exposed to component cause F, sufficient cause three would be formed in half of the people for whom the sufficient cause otherwise included D (half since half of those people – sufficient cause 2 – is not exposed to component cause B). In this example we can see that only 1/3 of the diseased can be attributed to component cause D (known as the *attributable fraction*), even though it is present in 2/3 sufficient causes (known as the *aetiologic fraction*). Please notice that this observation can be at odds with the interpretation of the *condicio-sine-qua-non* test that is applied in different judicial systems, for this principle does not necessarily provide the right mind-set to handle the possibility that a different causal mechanism leading to the same consequence could arise.

Although neither the counterfactual nor the sufficient cause of an individual can be observed, this conceptual framework does provide useful insight in the idea of causation and multi-causality.

V. Study Designs

The counterfactual ideal can be approached in several study designs, as long as several assumptions are made. Although uncommon, sometimes the counterfactual is undisputed and direct causal inferences can be made. For example, certain forms of brain injury can induce massive swelling of the brain which leads to increased intracranial pressure and subsequently the death of almost all patients with this condition.[10] Any intervention that reduces the intracranial pressure and prevents death in all patients, for example by drilling a hole in the skull so that the swollen brain can extent outward, will be re-

10 Zuurbier, S.M. *et al.*, 'Decompressive Hemicraniectomy in Severe Cerebral Venous Thrombosis: a Prospective Case Series', *Journal of. Neurology* 259 (2012), pp. 1099-105.

garded as causal in the prevention of death of these patients.

There will hardly be any discussion about the causal claim made in such a scenario, so we will not focus on this type of studies. We will focus on scenarios which are much more unclear. Since most, if not all, diseases can be regarded as multi-causal, the composition of sufficient cause of individual patients cannot be known, making it impossible to determine causal mechanisms in individuals. We can only quantify the effect of component causes in probabilistic terms.[11] Often this is done by comparing the risk of those who are exposed to the factor of interest to the risk of those who are not exposed, for example by the ratio of the respective probabilities of disease. This ratio is also known as the *relative risk*.

The study design that approaches the counterfactual ideal as close as possible is the crossover trial. In this design patients are assigned to two subsequent treatment strategies, of which one can be a placebo treatment, and the outcome of the patient (e.g. blood pressure) is measured directly after each treatment (e.g. antihypertensive medication vs. placebo). This way the same patient is observed both with and without the exposure, as prescribed by the counterfactual ideal. It is important that the patient has to return to his 'original state' from before his first treatment, before receiving his second treatment. Otherwise such a comparison will not result in correct causal inference. This problem can be countered by tweaking the experimental design, for example introducing a wash-out period between the two treatment periods, but also severely limits the applicability of this design.[12] Another study design that approaches the counterfactual ideal is the case-crossover design. In this design the exposure status of a patient is determined on two moments: acutely before the onset of the disease and in a control period some time before the onset of the disease. If the exposure of interest is indeed a cause of the disease it is likely to be more present just before the acute onset of the disease than in the control period. This can only be done when the information needed to determine exposure status can be reliably obtained after the patients are identified. Another disadvantage of this design is that it can only investigate triggers of diseases with an acute onset, which are the component causes with no or little induction time. An example of this study design is a study that investigated potential triggers of sub arachnoid bleeding, which showed that short but distinctive exposures such as coffee consumption and sexual intercourse can indeed be the trigger of this type of haemorrhagic stroke.[13,14]

Although these two study designs approach the counterfactual ideal, these can only be applied to situations in which an exposure is variable within one person and the effect is either acute or reversible. Many research questions do not adhere to these conditions (e.g. genetic exposures are not variable within a person, cancer has no acute onset and death is not reversible) thus leaving one or both of these crossover designs inappropriate. Other study designs do not suffer from these restrictions, but need more assumptions to justify causal inferences. Randomized trials can be used to study the effect of different treatment strategies by applying the treatments to different groups of persons and observe whether there is a difference in the frequency of the outcome of interest. This study design relies heavily on the assumption that the two groups would have a similar risk of the outcome if these were left untreated, a situation in which the counterfactual ideal clearly resonates. This situation is created by the randomization principle: the likelihood of receiving a certain treatment is independent from other causes of the outcome. Randomized trials are a powerful tool in the discovery of intended effects of modifiable exposures, being treatments targeted at reducing the risk of the outcome, as is the case in a clinical trial that compares two treatments to prevent cardiovascular disease. Also, data from randomized trials can provide more insight in the side effects of new drugs.

However, the use of randomized trials to identify causes of a disease is in many cases ethically undesirable. Additionally, many exposures cannot be modified (e.g. genetic variations) and therefore a large proportion of causal questions cannot be answered by experimental studies. In such cases observational studies must be applied to estimate the causal relationship between the exposure and the

11 Rothman, K.J., Greenland, S., Lash, T.L., *Modern Epidemiology* (third revised edition), Lippincott Williams & Wilkins 2008.

12 Senn, S., *Crossover-trials in clinical research*, Wiley 1993.

13 Maclure, M. *et al*, 'Should we use a case-crossover design?', *Annual Review of Public Health* (2000), pp. 193-221.

14 Vlak, M.H.M. *et al.*, 'Trigger Factors and Their Attributable Risk for Rupture of Intracranial Aneurysms: a Case-crossover Study', *Stroke* 42 (2011), pp. 1878–82.

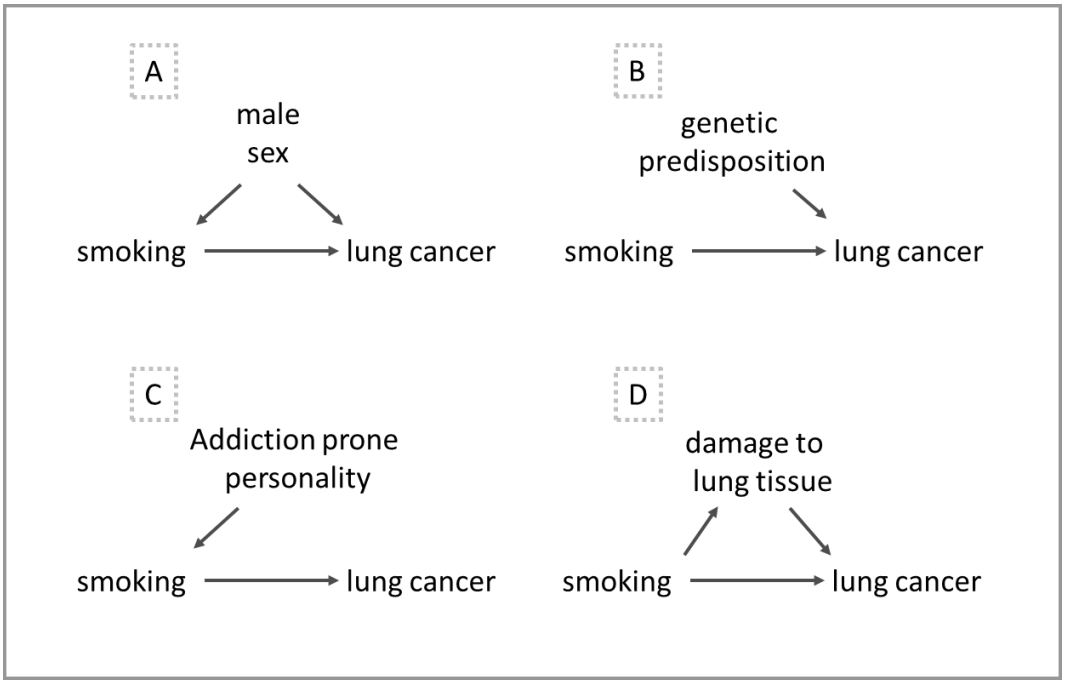

Figure 2

outcome of interest. The observational study designs can be categorised in two groups, being the cohort studies and the case-control study, each with their own merits. Like experimental study designs, observational study designs rely on certain assumptions to allow estimation of the causal effect. These designs, their merits and pitfalls as well as the assumptions needed for causal inference are too complex to describe here in detail and are discussed at great length in several textbooks and we limit ourselves to a general description of the concept of bias.[15]

VI. Bias

One major assumption in causal inference from epidemiological studies is the absence of bias, which introduces an incomparability into the study. We will discuss three major forms of bias with regard to the causal relationship between smoking and lung cancer. The first is *information bias* in which data are collected incorrectly and bias the result in a particular direction. For example when data about smoking habits are collected in a different fashion (for example more rigorously or through different types of questionnaires) in lung cancer patients than in healthy subjects. A comparison of those data would not only reflect the effect of smoking on the risk of developing lung cancer, but undesirably also reflects the differences in data collection. Another form of bias is *selection bias* in which study participation is dependent on the exposure and/or the outcome. For example, when lung cancer patients are compared to a group of healthy volunteers who are not reflective of the population from which the lung cancer patients arose, but are instead (indirectly) selected for being non-smokers, results of the comparison of these groups would not reflect the effect of smoking on the risk of developing lung cancer. It will undesirably be reflective of the differences between the two separate populations from which the patients and control group were sampled. A third form of bias is *confounding bias* in which the increase in risk of the exposure of interest is mixed with the risk of another cause of the disease of interest. This happens when the exposure of interest shares a common cause with the outcome of interest, as is discussed in figure 2.

15 Rothman, K.J., Greenland, S., Lash, T.L., *Modern Epidemiology* (third revised edition), Lippincott Williams & Wilkins 2008.

This figure contains four graphs that describe the causal relationship between smoking and lung cancer, but also include a third factor. These graphs are examples of four different classes of factors that are statistically associated to the risk of lung cancer, which could impede causal inference. It is important to differentiate between these classes because the nature of such a variable determines whether it should be taken into account to ensure valid estimation of the causal effect between smoking and lung cancer development.

A| A common cause of the exposure and outcome is considered a confounder. This example shows that men are more likely to smoke, but also that men intrinsically have a higher risk of lung cancer. The smoking-lung cancer association is said to be confounded and 'male sex' needs to be taken into account in order to ensure valid causal inference. Confounding can be a source of fallacious 'post hoc ergo propter hoc' conclusions.

B| Another cause of lung cancer, e.g. a genetic predisposition, which is independent of smoking is not considered a confounder. Therefore, the additional risk of some individuals will not confound the smoking-lung cancer association.

C| Causes of the exposure which are not a cause of the outcome other than via the exposure of interest are not confounders. In this example, an addiction prone personality is a causal factor in the development of a smoking habit. However, it is not a cause of lung cancer by itself. These causes are part of the causal mechanism of lung cancer, but do not confound the smoking-lung cancer association.

D| A direct consequence of the exposure which ultimately leads to the outcome of interest is not considered to be a confounder. In this example, smoking increases the risk of lung cancer because it causes damage to lung tissue. This intermediate cause is said to lie 'in the causal pathway'. Therefore, there is no confounding present.

The presence of confounding can lead to a fallacious 'post hoc ergo propter hoc' conclusion: even when an exposure of interest is not a cause of the disease, it is still possible that exposed individuals are more likely to develop the disease. This increase in risk, which is in fact a spurious relationship, can then be explained by other causes of disease that are found more often amongst exposed individuals leading to *confounding bias*. When confounding is not taken into account the disease develops more often in those with a certain exposure, it seems as if the exposure is in fact the cause. If sources of confounding are identified before the start of the study, confounding can be addressed and accounted for in the study design or statistical analyses.[16] However, when confounding is not sufficiently addressed, its presence may lead to erroneous causal statements.

All study designs are subject to bias, but different study designs suffer from different forms of bias and to a different extent. There are some classifications that categorise studies according to their 'level of evidence'.[17] This practice can be useful, as long as this practice does not preclude critical thinking. For example, many researchers believe that the randomized clinical trial is the only study design in which causal relationships can be studied. This is however an outdated point of view, since observational studies can be as credible as randomized trials under certain conditions.[18] The randomized controlled trial study design remains however the unbeatable golden standard if one wants to study the beneficial effects of a new drug. The randomization procedure breaks the link between the prescription of the new drug and the probability of the outcome. Observational studies do not break this link, which could severely bias the results (i.e. confounding by indication). However, these biases are less severe when one wants to study drug side effects or identify causes of a disease. This makes observational studies suitable to investigate causal mechanisms, in case biases can be accounted for.

VII. Causal Inference: More than One Study

So can we draw definite conclusions on the probabilistic relationship of a cause and its consequence based on a single study? It is advisable to use multiple studies for several reasons. First, it is possible that

16 For more background information on the statistical approaches that can be applied to investigate causal relationships, please refer to Berzuini, C., Dawid, S., Bernadinell, L., (editors), *'Causality: Statistical Perspectives and Applications',* (Wiley, 2012)

17 See for example the website of the Centre for evidence based medicine with the title 'Levels of evidence', <http://www.cebm.net/index.aspx?o=1025> (21 July 2014).

18 Vandenbroucke, J.P., 'When are observational studies as credible as randomised trials?', *Lancet* 363 (2004), pp. 1728–31.

just by chance the effect estimate from a single study is very different from the true effect. By combining the result of multiple studies into a so-called 'meta-analysis' the statistical power increases and the effect estimate is more precise. Second, all studies are subject to bias and some studies are more prone to particular forms of bias. Therefore, a lot can be learned from comparing the results of studies with different study designs. But even in the unlikely scenario that bias is thought to be completely absent and that the effect of the presumed cause is measured with sufficient statistical power, more information is needed to draw firm inferences on the causal relationship between the exposure and outcome of interest. This knowledge must focus on the plausibility of the proposed causal claim. Are other plausible factors present that could explain our results? Is the proposed mechanism in line with our current knowledge?

Therefore, part of causal inference in medicine lies outside the reach of a single study or even outside the realm of epidemiology. This concept is in line with the crossword analogy of science philosopher Susan Haack.[19,20] Several factors are of importance when filling out a crossword: the clue, the already entered answers, the possibility of alternative answers, and the level of completion of the crossword. A new answer cannot be at odds with already existing entries without rethinking previous answers. Causal inference can be regarded in a similar fashion: one single result is not likely to justify causal claims. But several results, from various research groups, backed by previous knowledge, not likely to be explained by alternative scenarios such as bias or chance could justify cautious causal claims about the quantification of the cause and effect estimate of interest.

Some have tried to codify all aspects that need to be considered before a relation can be regarded as causal. For example, Sir Austin Bradford Hill noted nine aspects of causality that might be considered when talking about causality in epidemiology.[21] Hill noted in his original address to the Royal Society that these factors are not to be considered as criteria. Only one, 'temporality', is a true criterion, that is that the cause must be present or act before its effect. The other eight aspects are not criteria and can be regarded as aspects that might be discussed when one wants to come to a causal judgement. However, despite the warnings by Hill and others, some researchers have misused these nine conditions as a checklist for causal claims. Such practice prohibits a critical appraisal of all evidence and should be abandoned. Unfortunately, this not the case.[22, 23]

VIII. Causal Claims in Law

It is easy to see that it is not straightforward to transfer epidemiological knowledge obtained from populations to individual legal claims. We will discuss these difficulties by discussing the Dutch Nefalit-case.[24] In this case, Karamus attributed his disease to his long-term exposure to asbestos, suffered in the factory where he worked, for which he held his former employer Nefalit liable. Nefalit had failed to take the necessary precautionary measures and was therefore, in the view of Karamus, to compensate all damages related to his disease. Nefalit responded, however, that the lung cancer could also have been caused by Karamus' long time smoking habit, by other factors or a combination of these. It is indeed known from epidemiological evidence as well as laboratory studies that both exposures are known to increase the risk of this particular type of lung cancer, often in combination with others causes. Therefore it is not possible, given the state of science and the idea of multi-causality, to determine the single cause of Karamus' disease and his damages. Lower courts, with the consent of the Dutch Supreme Court, acknowledged that applying the *condicio-sine-qua-non* test would mean passing on this uncertainty to Karamus entirely, as his claim would have to be dismissed on the ground that causation could not be estab-

19 Haack, S., *Manifesto of a Passionate Moderate*, Chicago: University of Chicago Press 1998.

20 Vandenbroucke, J.P., 'Alternative Medicine: A "Mirror Image" for Scientific Reasoning in Conventional Medicine', *Annals of Internal Medicine* 135 (2001), pp. 507-511.

21 Hill refers to these nine points as 'aspects of …(an) association' that should be considered before deciding on the interpretation of causation. These points are: strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment and analogy. See also Hill, A.B., 'The Environment and Disease: Association or Causation?', *Proceedings of the Royal Society of Medicine* (1965), pp. 295-300.

22 Morabia, A., 'On the Origin of Hill's Causal Criteria', *Epidemiology* 2 (1991), pp. 367-369.

23 Phillips, C.V. *et al.*, 'The Missed Lessons of Sir Austin Bradford Hill', *Epidemiology Perspectives and Innovations* 1 (2004), p. 3.

24 Hoge Raad 31 March 2006, ECLI:NL:HR:2006:AU6092, reachable through <http://uitspraken.rechtspraak.nl/#ljn/AU6092> (in Dutch; 9 December 2014).

Equation 1

$$attributable\ fraction\ (AF) = \frac{Relative\ Risk - 1}{Relative\ Risk}$$

lished. Therefore these courts applied the concept of so-called proportional liability, ruling that Nefalit was liable for only a proportion of Karamus' damages, based on expert testimony and epidemiological publications about the chances that his lung cancer was indeed caused by the asbestos exposure (55%).[25]

It was a matter of fairness, the Supreme Court indicated, not to pass on this uncertainty to the claimant entirely, by dismissing Karamus' claim altogether, given that in this case the chance that the lung cancer was indeed caused by asbestos, was neither very small nor very large. In such cases, courts are allowed to make a reasoned estimate, if necessary on the basis of expert testimony. It is important to note that the Supreme Court justified the application of this so-called 'proportional liability' in part by stating that there was uncertainty whether it was the asbestos exposure, the claimant's smoking habits, genetics or other additional external factors that caused the lung cancer, alone or in combination.

We will discuss later whether the 55%-ruling is justified in light of this motivation given by the Dutch Supreme Court. First, it is important to understand how the 55% came about. This number was obtained by calculating the *attributable fraction*, as discussed in section IV, which is defined as the fraction of cases in which the exposure of interest is a component cause of the sufficient cause leading to the disease. A second related measure is the *probability of causation*, which is a direct function of another fraction: the *aetiological fraction*. This fraction describes the probability that the factor of interest is a component cause in a sufficient cause, in a case randomly drawn from a patient population. In theory, these concepts can be very helpful in liability cases, because they provide a way to link a population measure to a single case. However, we have already argued that the aetiological fraction cannot be observed directly or calculated without strong additional assumptions, which cannot be empirically verified.

However, the *attributable fraction*, the fraction of the diseases among the exposed that can be ascribed to the exposure of interest, on the contrary can be calculated in a cohort study as (see Equation 1), where the relative risk is the risk of the outcome amongst the exposed divided by the risk in the unexposed. Once calculated the attributable fraction should directly be interpreted as the aetiological fraction: the aetiologic fraction is always similar or higher, but never lower than the attributable fraction.[26]

Some points have to be emphasized to ensure correct interpretation of these numbers. Both the aetiologic and attributable fraction are calculated for component causes, which implies that the sum of all fractions do not necessarily equal, but is likely to be higher than 100%, due to the multi-causal nature of complex diseases. In fact, the sum of these fractions could be both higher or lower, and basically depends largely on the number of causes that have been identified for a specific disease. Therefore, these fractions should never be interpreted as the probability that a certain factor of interest is *the single cause* of the disease in a particular case, since there is no such thing as a *single cause*. Some have proposed this wrong definition in order to use the effect size as a measure of causality. In line with this wrong notion a relative risk greater than 2, which equals an attributable fraction of > 50% (AF = (2-1) / 2), has sometimes even been abused as cut off point for 'causality-proven vs.

25 Hoge Raad 31 maart 2006, ECLI:NL:HR:2006:AU6092. See also, more recently, Hoge Raad 14 december 2012, ECLI:NL:HR:2012:BX8349. On these cases, see Castermans, A.G. & Hollander, P.W. den, 'Omgaan met onzekerheid. Proportionele aansprakelijkheid, artikel 6:101 BW en de leer van de kanss-chade', *NTBR* 2013, pp. 185-195 (in Dutch).

26 The situation under which the attributable fraction can be interpreted as the aetiological fraction are described in Kenneth J. Rothman, Sander Greenland, Timothy L. Lash. *Modern Epidemiology,* third revised edition, (Lippincott Williams & Wilkins, 2008)

Equation 2

$$aetiological\ fraction = \frac{number\ of\ \textbf{sufficient causes}\ that\ include\ the\ component\ cause\ of\ interest}{all\ possible\ \textbf{sufficient causes}\ leading\ to\ this\ disease}$$

Equation 3

$$proportional\ liability = \frac{number\ of\ \textbf{component causes}\ that\ are\ the\ responsibility\ of\ the\ defendant}{total\ number\ of\ \textbf{component causes}\ in\ the\ sufficient\ cause\ of\ the\ claimant}$$

causality not proven'.[27] This misuse of the attributable fraction precludes any form of critical thinking about the causal mechanism underlying events and should be abandoned.

Another possible misinterpretation of both the aetiologic and attributable fraction lies in the direct translation of the attributable fraction to the proportion of the claims that should be reimbursed, with the idea that on average both the plaintiff as well as the defendants are treated satisfactorily. However, by coupling the attributable fraction to the proportion that should be reimbursed, the court forgets a crucial characteristic of the attributable risk, which again is that the sum of the attributable fraction can exceed 100%. In contrast, the shares in proportional liability in one particular case should not. Consider again our example in figure 1, in which 100% of cases (3/3) was 'caused by A' and 66% of all cases (2/3) was 'caused by B'. If a claimant with this particular disease would theoretically hold both 'A' and 'B' liable in separate law suits, this approach would yield a total of 166% of the claimed sum, which does not adhere to the fairness principle. The misconception that the aetiologic or attributable fraction can directly be applied as an allocation instrument for proportional liability as a legal concept thus lies in erroneously applying a population measure to an individual probability estimation. This can also be appreciated when we compare the formula for the aetiological fraction (see Equation 2) to the concept that uses proportional liability to adhere to the fairness principle (see Equation 3).

So what to think then of the use of proportional liability in the case of Nefalit and Karamus? During the hearings, an expert motivated that there was a 125% increase in risk due to asbestos exposure, which corresponds to a relative risk of 2.25 and an attributable fraction of 55% (the AF = (2.25-1)/2.25= 55.56%, the lower court mentions 55% in its ruling). The Dutch Supreme Court motivated the use of proportional liability, including this figure, and thereby implicitly the use of the attributable fraction in its ruling with the observation that there was uncertainty whether asbestos was indeed the cause. However, the court went further by coupling this number as the fraction of the damages that employer Nefalit should reimburse as a matter of fairness. At first glance, the motivation of the Supreme Court sounds fair, but we have already showed in our example above that linking the attributable fraction to the fraction that should be reimbursed by the defendant does not always adhere to the matter of fairness. Therefore, the ruling by the Supreme Court could lead to unfair reimbursements and, perhaps unknowingly and unwantedly, sets a precedent with possibly unwanted consequences.

We will continue with the Nefalit-case to illustrate this. Let say that besides smoking and asbestos exposure the claimant was also subjected to another risk factor 'X' due to negligence of another employer. Again, it is uncertain whether indeed it was 'X' that was the cause of his disease. Let us state that 'X' increases the risk of lung cancer by 178% and therefore has an attributable fraction of 64% (i.e. a relative risk = 2.78 and AF = (2.78-1)/2.78). Following the same line of reasoning as the court did when it came to asbestos exposure (i.e. there is uncertainty about the causal claim and therefore only a part of the claim should be reimbursed), in theory 64% of the claim

27 Greenland, S., 'Relation of Probability of Causation to Relative Risk and Doubling Dose: a Methodologic Error That Has Become a Social Problem', *American Journal of Public Health* 89 (1999), pp. 1166–9.

should be reimbursed by the second employer. This makes the received amount to theoretically supersede the original claim.

When a court wants to directly couple the aetiological faction to a 'fair' distribution of the damages the court has to know the true underlying causal mechanism of each individual liability claim. In a sense, the court has to be certain about all the component causes that make up the sufficient cause in this particular individual. However, the exact sufficient cause cannot be observed in an individual case, an uncertainty that the Supreme Court used to motivate its ruling. So, when a court is willing to assume proportional liability, it should be well motivated. Even more, when a court is uncertain whether the defendant is indeed responsible for one of the component causes in this particular case, it is even more difficult to understand how it can be justified to link the proportional liability to the aetiological fraction, its derivatives and approximations.

Based on these points, it is already highly questionable whether proportional liability should be directly linked to epidemiological population measures such as the unobservable aetiological fraction or the attributable fraction as its derivative. But the most important objection of this direct coupling is the fact that the sum of these numbers are not restricted to, and is even very likely to supersede, 100%. We do see the merit of proportional liability, especially given the multi-causal nature of most diseases, and we would therefore like to propose a different approach that links these two concepts without the aforementioned problems. For this, we will use the component cause concept in combination with the *condicio-sine-qua-non*-principle in a two-stages approach.

IX. Proportional Liability in Two Stages

The approach we would like to propose is a two-stages-approach, linking the concepts of proportional liability and multi-causality. This approach makes use of the *condicio-sine-qua-non*-test and thus provides equal weights to all possible causes. This is in line with the notion of both the sufficient cause model and the counterfactual model.

During the first stage of this approach, the court has to decide whether the defendant's wrongful behaviour indeed played a role in the causal mechanism. The court should motivate its decision on evidence and expert witnesses. Once decided whether the defendant indeed played a role in the causal mechanism (i.e. is responsible for one or more component causes of the sufficient cause), the defendant can advocate proportional liability in the second stage. The defendant does so by providing a list of possible other component causes to the court, of which it has to determine whether these also played a role in this particular case. This way, the court can determine the fraction of component causes part of the presumed sufficient cause, that are the responsibility of the defendant. This fraction could be used to determine proportional liability (cf. equation 2). For example, when there are six possible causes, of which four might play a role in the case at hand and one of these four can be attributed to the defendant, the defendant would have to compensate 25% of the claim.

This two-stages-approach is not flawless, for it could overestimate the number of component causes that play a role in the sufficient cause and thereby underestimate the liability of the defendant. Also, new component causes could be identified after the court has decided. If this would lead to a new liability claim with a new defendant, our example could be summarised as follows. With the discovery of a new cause that is relevant to our case, there are now seven component causes of which five are applicable to the case at hand. If one of those component causes can be attributed to the second defendant, then he would have to pay 20% of the original claim. This way, the total sum of all claims, 40% in our example, will never supersede 100% of the original claim, but approaches this number asymptotically. Receiving this 20% of the second defendant should be conditional on the reimbursement of the excess 5% that was paid by the first defendant.

Another problematic aspect of this two-stage-method is that all possible component causes are considered equally important and are given the same weight in this approach. Although this is in line with the component cause model, it does result in some practical problems. For example, there can be numerous component causes which might be listed that indeed are component causes in the most strict definition, but lack relevance when it comes to proportional liability (e.g. one has to have lungs in order to develop lung cancer). Also, evidence might suggest that some component causes cannot be discarded, but are certainly less relevant to the case in question then others. In that case, a weighted approach could be

Box 1 - Take home messages

- Causal claims should always be considered in the light of multi-causality: there is never *the* cause, but a set of component causes that make up a sufficient cause.
- Causality in epidemiology relies on more than just one study: different studies, the effect of possible biases and additional evidence, even outside the realm of epidemiology, should all be taken into account before cautious claims can be made.
- The aetiological fraction and the probability of causation as its derivative are both epidemiological measures which cannot be calculated. They can only be approached, under certain assumptions, by calculating the attributable fraction.
- Linking the concept of proportional liability to the attributable fraction is wrong, especially because the sum of all attributable fractions is likely to exceed 100%.

considered. All in all, it is up to the court, with the aid of experts and scientists, to rule which possible component causes are relevant to the question of liability.

X. Conclusion

Causality research in epidemiology is largely embedded in the concept of the counterfactual model, which resembles the legal *condicio-sine-qua-non*-test. By definition, the counterfactual cannot be observed and the sufficient cause in a single person cannot be known. Therefore, it is not possible to know the exact causal mechanism leading to the disease in an individual person. However, epidemiological studies can be used to study the effect of a presumed cause on the risk of disease at the population level. Results from multiple and reliable studies, considering multi-causality, combined with prior biological knowledge can result in cautious causal claims. Although the aetiologic fraction can never be known, the attributable fraction can be calculated and gives insight in the relation between cause and effect on a group level.

This population measure cannot directly be applied to individual cases without relying on untestable assumptions (see Box 1).

Linking the concept of proportional liability to the attributable fraction is thus wrong. In addition, the sum of the attributable aetiological fractions is likely to exceed 100%, which could lead to unfair reimbursements. We have therefore proposed a two-stage-approach for a court to apply the concept of proportional liability, by first deciding on liability and then on the proportion. This links proportional liability to the concept of multi-causality, while also and firstly adhering to the *condicio-sine-qua-non*-test. In this process, the court should consult scientist and experts, but ultimately, the decision remains a normative judgment for the court itself to make.